

基于近红外光谱的北方潮土土壤参数实时分析

孙建英¹, 李民赞^{1*}, 郑立华¹, 胡永光^{1,2}, 张喜杰¹

1. “现代精细农业系统集成研究”教育部重点实验室, 中国农业大学, 北京 100083

2. 江苏大学机械工程学院, 江苏 镇江 212013

摘要 选取中国北方潮土作为研究对象, 探索利用近红外光谱分析技术分析土壤参数的可行性和可能性。从一块试验麦田共采集了150个土样, 土样在采集回实验室后, 在保持其原始状态条件下利用傅里叶变换近红外光谱仪迅速测定了其近红外光谱。近红外光谱变量为原始吸收光谱和一阶微分光谱, 分析的土壤参数有土壤水分、有机质和全氮的含量。对于土壤水分, 在相关分析的基础上建立了一元线性模型, 所采用的波长为1 920 nm, 模型的相关系数达到0.937, 模型可以直接用于土壤水分的实时预测。对于有机质和全氮含量建立了多元回归模型, 有机质预测模型所采用的波长是1 870和1 378 nm, 全氮预测模型所采用的波长则是2 262和1 888 nm。分析结果表明土壤有机质和全氮含量可以利用田间土样的近红外光谱特性进行分析和检测, 建立的线性模型是有效的。

关键词 近红外光谱; 化学计量学; 土壤水分值; 土壤有机质含量; 土壤全氮含量

中图分类号: TH83; S15 **文献标识码**: A **文章编号**: 1000-0593(2006)03-0426-04

近红外光谱分析方法是一种快速、简便及无破坏性的测量方法^[1]。有很多研究介绍利用此方法分析土壤参数, Hummel等分析土壤水分和有机质与近红外光谱之间的关系, 利用逐步多元回归的方法建立了预测模型^[2]。Krishnan等分析土壤有机质和近红外光谱之间的关系, 利用4个波长(1 130, 1 350, 1 398, 2 210 nm)进行多元线性回归分析建立模型, 相关系数达到0.934^[3, 4]。彭玉魁等^[5]分析土壤水分、有机质和总氮含量与近红外光谱之间的关系, 利用多元回归方法建模得到的相关系数分别为0.969, 0.921和0.928。以上研究都是利用干燥后的土样, 或是在实验室经过处理后的土样, 其分析结论尚不能直接用于实时分析土壤的参数, 因此需要分析利用田间原始状态土壤的近红外光谱分析土壤参数的可能性和可行性。另外, 研究表明土壤类型对土壤光谱有着较大的影响, 而北方潮土又是我国北方粮食生产区的主要土壤类型, 因此以我国北方潮土为研究对象探索基于近红外光谱的土壤参数预测也是非常必要的。本研究即是以我国北方潮土作为研究对象, 通过对田间原始土样的近红外光谱进行分析, 建立了土壤参数的统计预测模型。

1 材料和方法

土样采集于位于北京郊区的中国农业大学一块试验麦

田, 采集的土样直接用于近红外光谱和水分的测量, 近红外光谱采用由美国 Thermo Nicolet 公司生产的 Antaris 型傅里叶变换近红外光谱仪测量, 它配有光导纤维探头, 可以直接将探头插入土中进行测量。光谱仪的光谱测量范围为12 000~4 000 cm^{-1} (833~2 500 nm); 分辨率为2 cm^{-1} ; 波数精度: 0.1 cm^{-1} 。扫描次数: 32。

完成光谱测量的土样被分为两部分, 一部分直接用来分析土壤水分, 另一部分风干后用于分析土壤肥力成分。水分的测量是利用电热恒温干燥箱, 保持105 $^{\circ}\text{C}$, 放置24 h进行烘干后得到土壤的水分含量。风干后的土样进行化学分析, 分析的成分或参数有土壤有机质含量(SOM), 土壤全氮含量(TN), 土壤全磷含量, 土壤全钾含量以及土壤pH值等。作为研究的第一步, 本研究以三个最重要的土壤参数, 即土壤水分、SOM和TN为研究对象, 探索了基于近红外光谱的北方潮土土壤参数的实时分析。

2 结果和讨论

2.1 光谱测定结果和土壤参数分析结果

图1给出了光谱测定结果的例子。分别选取在同一有机质和N的含量的基础上, 水分含量相差较大的两个土样(1和4)和同一水分基础上有机质含量相差较大的两个土样光

收稿日期: 2005-03-03, 修订日期: 2005-06-17

基金项目: 国家自然科学基金(30370812)资助项目

作者简介: 孙建英, 女, 1981年生, 中国农业大学“现代精细农业系统集成研究”教育部重点实验室硕士研究生 * 通讯联系人

谱(2 和 3)进行比较可以看到:(1)由水分所引起的三个吸收峰在每个光谱曲线中都清晰可见,这三个吸收峰分别大约是 1 450, 1 940 和 2 100 nm;(2)由图 1 得到,水分不同时不仅在水分的吸收峰处,在整个近红外区域吸光度都呈现出较大的差别,随着水分的增加,吸光度也明显增加;(3)由图 1 得到,当水分接近时,虽然土壤肥力参数有较明显差异,但吸光度光谱并没有体现出相应的变化,特别是在本实验中土样采集于田间,各种肥力成分没经过调制,相互之间差别较小,反映在光谱上的差别会更小。(4)由以上几条可以得出,土壤的近红外光谱受土壤水分的影响更大些,而受 SOM, TN 这些肥力参数的影响要小得多,因此需要新的数据处理和建模方法建立这些参数的预测模型。

表 1 给出了土壤参数测定结果的统计特性^[5]。水分值的变化较大,具有较大的方差,便于建立预测模型;而 SOM 和 TN 的含量基本上都在均值附近,标准差也较小,如前所述,这会给建模带来一定的困难。

虽然利用吸光度光谱数据可以进行相关及回归分析,由于一阶微分光谱会消除一些系统误差,提高建模精度^[6],本研究对每个吸光度光谱也都计算了一阶微分。

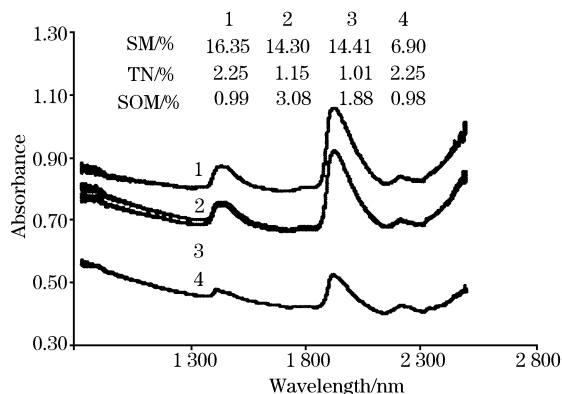


Fig. 1 NIR absorbance spectra of soil sample

Table 1 The measured values of three parameters of soil in the laboratory

参数	样品数	含量/%	均值/%	标准差/%
水分值(SM)	150	6.901~18.552	13.863	2.471
全氮	150	0.12~1.68	1.060	0.142
有机质	150	1.63~3.14	2.398	0.351

2.2 水分的分析

土样的原始吸收光谱与水分值有很大的相关性,将 150 个近红外光谱数据随机分为两个集合:标定集合和验证集合^[7]。其中标定集合有 90 个样本,用来建立预测模型;验证集合有 60 个样本,用来检验预测模型。首先计算了在整个近红外区土壤水分与吸光度光谱之间的相关系数,所有的值都高于 0.93。由于相关系数较高,决定建立模型稳定性较好的一元线性模型。

选取吸光系数最高的峰值点 1 920 nm 处的吸光度值与水分值做一元线性回归(SLR)分析。建立的一元线性回归模

型如下,

$$y = 17.706 - 2.167x_{1.920} \quad (1)$$

这里 y 为预测水分值, x_i 为原始吸收光谱数据。

建立模型的相关系数为 $r^2 = 0.936$, F 检验和 t 检验都显示模型以及各个模型参数显著。随后利用验证集合的 60 个数据对模型进行验证,标定及验证的结果如图 2 所示。模型验证的相关系数为 $r^2 = 0.935$ 。

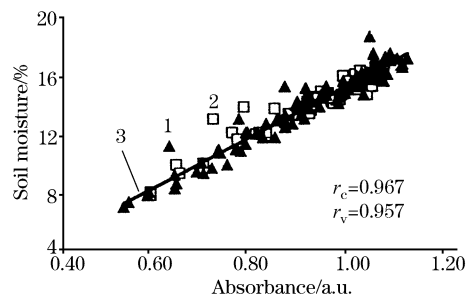


Fig. 2 Calibration and validation of soil moisture SLR model

1, Calibration; 2, Validation; 3, Model

2.3 全氮的分析

首先也计算了每个波长的吸收光谱与土壤全氮的相关系数,结果显示在 800~2 500 nm 波段内,土壤全氮含量与原始吸收光谱之间有较低的相关性,最大不超过 0.4。分析原因可以认为主要由以下两点:(1)含氮化合物的吸光系数较低,含氮化合物的基准振动吸收频率多在中红外区段,其波长范围约在 5 000~10 000 nm。近红外区是这些基准振动的第二、第三倍频或组合频,能级跃迁发生的概率比基准振动要低得多,因此造成吸光系数也较低。(2)由于土壤样品是采集于农田的实际土壤,土壤的含氮量没有经过调制,使得得到的土样的 TN 含量的标准差较低,从而影响了相关系数。但是根据近红外吸收理论,由于含氮化合物在近红外区有确定的吸收频带^[1],利用近红外光谱分析土壤全氮含量是可行的,因此决定寻找一种对现有田间原始土样的测定值进行预处理的方法,在预处理之后再分析吸光度光谱和 TN 之间的相关性,并建立预测模型^[8]。

2.3.1 光谱和 TN 含量数据的预处理

土壤全氮含量最大、最小的范围为 1.68%~0.12%,取间隔为 0.055 将 150 个数据分为 30 组,每一个组内的数据做平均后得到 30 个新的土壤全氮含量数据,对于每一个组内的土壤全氮含量所对应的近红外光谱数据也做平均,最后得到与土壤全氮含量相对应的 30 个近红外光谱数据,图 3 是重组后的两个特定波长处的光谱数据与土壤 TN 含量之间的散点图,图中显示吸光度与土壤全氮之间存在有较高的相关关系。对于得到新数据的每一个波长也计算相关系数,结果显示在 TN 的主要吸收频带,都获得了较高的相关系数。按照同样的预处理方法,还得到了 15 个用于模型验证的新数据。

2.3.2 偏最小二乘回归分析

选取相关系数比较大的四个波长 2 240, 2 090, 1 901 和 1 827 nm 进行了多元线性回归(MLR)分析,但结果显示光谱数据之间具有很强的多重共线性。因此,利用偏最小二乘回归(PLSR)对以上同样的四个波长进行分析。偏最小二乘

回归采用对数据信息进行分解和筛选的方式,有效提取对系统解释性最强的综合变量,剔除多重相关信息和无解释信息的干扰^[9]。

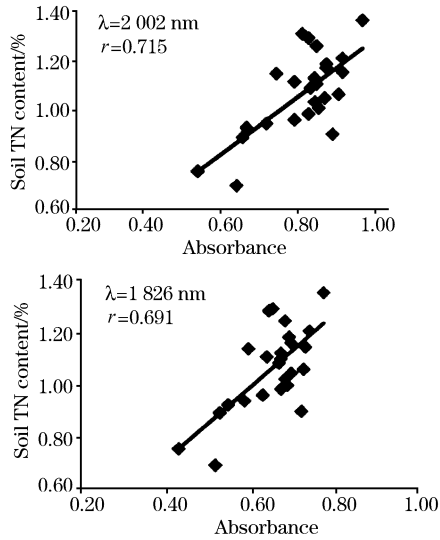


Fig. 3 Correlation between original absorbance and soil TN content

利用交叉有效性确定选取的成分数为两个,建立的回归模型(还原后的)如下,

$$y = 0.237 - 2.492x_{2240} + 0.263x_{2090} + 6.138x_{1901} - 4.285x_{1827} \quad (2)$$

这里 x_i 为第 i 个波长的光谱吸光度, y 为 TN 的含量。

模型标定和检验的结果见图 4(a), 预测模型的相关系数为 $r^2 = 0.601$, 模型检验的相关系数为 $r^2 = 0.643$ 。另一方面从图中看出, TN 含量和光谱数据之间还含有较强的非线性成分。

2.3.3 一阶微分光谱的多元回归分析

由于理论上微分光谱可以消除系统误差,也建立了基于微分光谱的 TN 含量预测模型。同样选取四个相关系数较大的波长(2262, 1888, 1408 和 1082 nm)的一阶微分光谱做多元线性回归分析。得到的模型通过了 F 检验,但是波长为 2262 和 1408 nm 的系数没有通过 t 检验。因此选取 1888 和 1082 nm 做多元回归分析,建立的回归模型为

$$y = 0.758 - 149.807x_{1082} - 51.828x_{1888} \quad (3)$$

这里 x_i 为一阶微分吸收光谱, y 为 TN 的含量值; 模型标定和模型检验的结果见图 4(b), 模型的相关系数为 $r^2 = 0.767$, 从点图的分布来看,非线性成分也已大大降低了。 F 检验值(Significance F) 为 1.069×10^{-7} , 接近于零,即 F 检验结果显著,各参数的 t 检验对应的 P -value 分别为 0.001, 0.019, 均小于 0.05, 即各参数 t 检验也都显著。

2.4 有机质的分析

土样的有机质含量的差异也很小,原始吸收光谱及一阶微分光谱与有机质之间的相关系数也偏低,最大不超过 0.4, 利用与 TN 同样的方法对有机质含量数据进行了预处理,利用得到的 30 个新数据进行建模。仍然利用一阶微分光谱作

为变量进行分析。选取相关系数较大的两个波长 1870 和 1038 nm 做多元线性回归分析,建立的模型如下

$$y = 1.195 - 515.47x_{1870} + 159.268x_{1038} \quad (4)$$

这里 x_i 为一阶微分吸收光谱, y 为有机质的含量; 建模结果如图 5 所示。

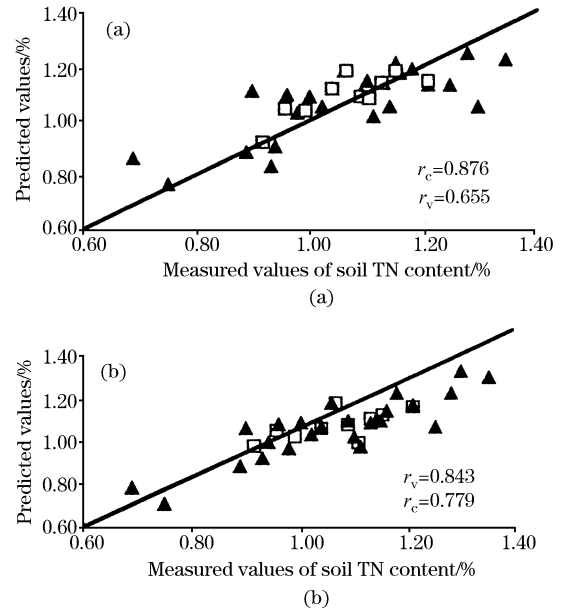


Fig. 4 Calibration and validation of TN content

(a), Calibration and validation of PLSR model; (b), Calibration and validation for the model of the first derivative

▲, Calibration; □, Validation

模型的相关系数为 $r^2 = 0.607$, F 检验值(Significance F) 为 3.423×10^{-6} , 接近于零,即模型的线性检验结果为显著。三个模型参数的 t 检验值对应的 P -value 分别为 0.0005, 0.0006 及 0.0037, 均小于 0.05, 由此可得各参数均通过了 t 检验。利用得到的 15 个新数据对模型进行验证,也获得了较高的相关系数, $r^2 = 0.607$ 。

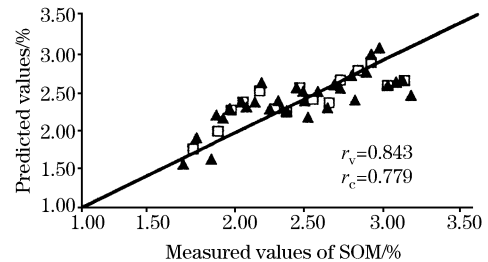


Fig. 5 Calibration and validation of SOM content

▲, Calibration; □, Validation

3 结论

利用傅里叶变换近红外光谱仪对采集的原始土样测量得到吸收光谱数据,然后利用相关分析及回归分析建立了土壤水分、全氮含量和有机质含量的预测模型,主要结论如下。

(1) 土壤水分值与原始吸收光谱有着很高的相关系数,

选取 1 920 nm 做一元线性回归分析, 得到了高精度的预测模型。由于选取的波长单一, 建立的模型简单, 适合用于作为开发便携式土壤水分测试仪的理论基础。

(2) 通过对预处理后的土壤全氮和土壤有机质含量数据的相关分析, 得到了具有较高精度的基于微分光谱的预测模型。结果显示基于原始土样的近红外光谱检测土壤全氮和土壤有机质含量是可行的。而且由于微分光谱能在一定程度上消除系统误差, 基于微分光谱的模型具有更高精度和适用性。

(3) 本研究以北方潮土田间原始土样的光谱作为基础,

建立了三种土壤参数的预测模型, 所得到的结果可以成为土壤实时分析的基础。为了进一步提高模型精度, 有必要在下一步的试验中, 对田间的土壤实施必要的养分胁迫或调控, 加大土样之间养分的变异, 从而建立更实用的土壤肥力预测模型。

致谢: 土壤肥力参数测量是由中国农业大学资环学院化学分析室负责完成的, 深表谢意。同时感谢中国农业大学精细农业研究中心赵朋、张俊宁、杨玮、孔德秀、崔笛、唐宁同学在采样和实验期间所给予的帮助。

参 考 文 献

- [1] LU Wan-zhen, YUAN Hong-fu, XU Guang-tong, et al(陆婉珍, 袁洪福, 徐广通, 等). Modern Analysis Technique of NIR(现代近红外光谱分析技术). Beijing: China Petrochemistry Press(北京: 中国石化出版社), 2000. 4.
- [2] Hummel J W, Sudduth K A, Hollinger S E. Computers and Electronics in Agriculture, 2001, 32: 149.
- [3] Krishnan P, Alexander J D, Butler B J, et al. Soil Sci. Soc. Am. J., 1980, 44: 1282.
- [4] Krishnan P, Butler B J, Hummel J. Transactions of the ASAE, 1981, 24(2): 306.
- [5] PENG Yu-kui, ZHANG Jian-xin, HE Xu-sheng, et al(彭玉魁, 张建新, 何绪生, 等). Acta Pedologica Sinica(土壤学报), 1998, 35(4): 553.
- [6] YAN Yan-lu, ZHAO Long-lian, LI Jun-hui, ZHANG Lu-da, MIN Shun-geng(严衍禄, 赵龙莲, 李军会, 张录达, 闵顺耕). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2000, 20(6): 777.
- [7] LI Ning, MIN Shun-geng, QIN Fang-li, ZHANG Ming-xiang, YE Sheng-feng(李 宁, 闵顺耕, 覃方丽, 张明祥, 叶升锋). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(1): 45.
- [8] GAO Rong-qiang, FAN Shi-fu, YAN Yan-lu, ZHAO Li-li(高荣强, 范世福, 严衍禄, 赵丽丽). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(12): 1563.
- [9] WANG Hui-wen(王惠文). Partial Least-Squares Regression-Method and Applications(偏最小二乘回归方法及应用). Beijing: National Deference Industry Press(北京: 国防工业出版社), 1999.

Real-Time Analysis of Soil Moisture, Soil Organic Matter, and Soil Total Nitrogen with NIR Spectra

SUN Jian-ying¹, LI Min-zan^{1*}, ZHENG Li-hua¹, HU Yong-guang^{1, 2}, ZHANG Xi-jie¹

1. Key Laboratory of MOE on Modern Precision Agriculture System Integration Research, China Agricultural University, Beijing 100083, China
2. School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China

Abstract The grey-brown alluvial soil in northern china was selected as research object, and the feasibility and possibility of real-time analyzing soil parameter with NIR spectroscopic techniques were explored. One hundred fifty samples were collected from a winter wheat farm. NIR absorbance spectra were rapidly measured under their original conditions by a Nicolet Antaris FT-NIR analyzer. Three soil parameters, namely soil moisture, SOM (soil organic mater) and TN (total nitrogen) content, were analyzed. For soil moisture content, a linear regression model was available, using 1 920 nm wavelength with correlation coefficient of 0.937, so that the results obtained could be directly used to real-time evaluate soil moisture. SOM content and TN content were estimated with a multiple linear regression model, 1 870 and 1 378 nm wavelengths were selected in the SOM estimate model, and 2 262 and 1 888 nm wavelengths were selected in the TN estimate model. The results showed that soil SOM and TN contents can be evaluated by using NIR absorbance spectra of soil samples.

Keywords NIR spectroscopy; Chemometrics; Soil moisture; Soil organic matter; Soil total nitrogen

* Corresponding author

(Received Mar. 3, 2005; accepted Jun. 17, 2005)